

University of Groningen

**Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness**

Schaafsma, Joanna D.; van der Graaf, Yolanda; Rinkel, Gabriel J. E.; Buskens, Erik

*Published in:*  
Journal of Clinical Epidemiology

*DOI:*  
[10.1016/j.jclinepi.2009.01.008](https://doi.org/10.1016/j.jclinepi.2009.01.008)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2009

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Schaafsma, J. D., van der Graaf, Y., Rinkel, G. J. E., & Buskens, E. (2009). Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. *Journal of Clinical Epidemiology*, 62(12), 1248-1252. <https://doi.org/10.1016/j.jclinepi.2009.01.008>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

SYSTEMATIC REVIEW

# Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness

Joanna D. Schaafsma<sup>a,\*</sup>, Yolanda van der Graaf<sup>b</sup>, Gabriel J.E. Rinkel<sup>a</sup>, Erik Buskens<sup>b,c</sup>

<sup>a</sup>Department of Neurology, University Medical Center Utrecht, The Netherlands

<sup>b</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands

<sup>c</sup>Department of Epidemiology, University Medical Center Groningen, The Netherlands

Accepted 16 January 2009

## Abstract

**Objective:** The lack of a standard methodology in diagnostic research impedes adequate evaluation before implementation of constantly developing diagnostic techniques. We discuss the methodology of diagnostic research and underscore the relevance of decision analysis in the process of evaluation of diagnostic tests.

**Study Design and Setting:** Overview and conceptual discussion.

**Results:** Diagnostic research requires a stepwise approach comprising assessment of test characteristics followed by evaluation of added value, clinical outcome, and cost-effectiveness. These multiple goals are generally incompatible with a randomized design. Decision-analytic models provide an important alternative through integration of the best available evidence. Thus, critical assessment of clinical value and efficient use of resources can be achieved.

**Conclusion:** Decision-analytic models should be considered part of the standard methodology in diagnostic research. They can serve as a valid alternative to diagnostic randomized clinical trials (RCTs). © 2009 Elsevier Inc. All rights reserved.

**Keywords:** Diagnostic research; Test characteristics; Clinical outcome; Cost-effectiveness; Diagnostic randomized clinical trial; Decision model

## 1. Introduction

To date, consensus on the methodology to evaluate new diagnostic tests is lacking [1]. Moreover, with rapid technical advances, especially in the field of imaging, diagnostic techniques undergoing evaluation may be already outdated before diagnostic and clinical values are established [2,3].

Similar to therapeutic research, a hierarchy can be discerned within diagnostic research [1,2,4–7]. The first step is the assessment of test characteristics of a new test. The next step is the evaluation of its added value [8]. The third step is the assessment of the effect on clinical outcome, and the final step comprises a cost-effectiveness analysis [7]. For evaluation of clinical outcome and cost-effectiveness of tests, randomized clinical trials (RCTs) are often not feasible.

We will shortly review each step of diagnostic research and discuss decision analysis as a useful alternative

methodology for critical assessment of clinical value and efficient use of resources.

## 2. The hierarchy in diagnostic research

We will illustrate the different phases of diagnostic research following the example of carotid artery stenosis, a well-known risk factor for stroke. Carotid endarterectomy reduces this risk in selected patients [9,10]. The standard technique for grading stenosis used to be digital subtraction angiography (DSA), an invasive imaging modality. Later on, magnetic resonance angiography (MRA) and duplex ultrasound (DUS) were introduced as noninvasive alternatives [11].

### 2.1. Step 1: Test characteristics

Test characteristics of a diagnostic tool provide information on the ability to discriminate between the absence and the presence of disease. They are expressed in terms of sensitivity, specificity, predictive values, and likelihood ratios. Sensitivity and specificity are useful in selecting tests, whereas predictive values provide information on the probability of disease, given a certain test result [1]. Likelihood ratios characterize the change in the probability of disease

\* Corresponding author. Department of Neurology, University Medical Center Utrecht, Heidelberglaan 100, PO Box 85500, 3508 GA Utrecht, The Netherlands. Tel.: +31-88-755-8600; fax: +31-30-254-2100.

E-mail address: j.d.schaafsma@umcutrecht.nl (J.D. Schaafsma).

**What is new?**

- In diagnostic research, decision models form a valuable alternative when randomized clinical trials (RCTs) are infeasible by integrating the best available evidence.
- Decision-analytic models should be increasingly used to evaluate clinical outcome including cost-effectiveness of fast developing diagnostic techniques.

after completing the test compared with the probability of disease before completing the test.

Test characteristics should be evaluated in a blinded cross-sectional study. Each test including a reference test needs to be performed in all study subjects to enable direct comparisons of the tests. Furthermore, the study participants ought to be representative of the target population [12,13]. In diagnostic tools that require interpretation, intra- and interobserver variability should also be assessed. In carotid artery stenosis, test characteristics of MRA and DUS have been established in a blinded cross-sectional study with DSA as the reference test [14].

A limitation of assessing test characteristics is dichotomization of test results, whereas only a few tests yield just “presence” or “absence” of disease as test results. For continuous and ordinal test results, a threshold needs to be established to determine whether a result is positive or negative. Ideally, this positivity criterion should reflect the clinical impact of false-positive and false-negative results. In case of carotid stenosis, the consequences of false-positive results, such as complications of unnecessary treatment, should be weighed against the consequences of missing severe stenosis. Dichotomization generally leads to loss of information. To avoid this, likelihood ratios can be used. They represent the ratio between the likelihood of a particular test result in patients with a certain disease status and the likelihood of the same test result in patients without this disease status. In case of multiple test results, the likelihood ratio can be calculated for each test result. Subsequently, the probability of the presence of disease, given a certain test result, can be calculated starting from the probability of disease before the test result was known and the likelihood ratio of the pertaining test result [15,16].

A further limitation of test characteristics is the use of a “gold” standard to establish a “true” disease status. Within this framework, a new test can never outperform its reference test, which in daily practice has frequently happened. For example, computed tomography (CT) scanning was clearly better than skull radiography for assessing intracranial pathology.

Moreover, test characteristics are not constant, but are influenced by factors, such as prevalence of disease, disease

severity, gender, age, and comorbidity. The higher the prevalence of disease, the higher the predictive value of a positive test result and the lower the predictive value of a negative test result. Apart from the predictive values, sensitivity and specificity may also be influenced by disease prevalence. However, the latter effect is indirect through prevalence varying with disease severity or disease spectrum. The probability of disease changes when information is obtained from the history, physical examination, and prior tests, because this information is used to select patients with a higher probability of disease for further diagnostic testing [17]. Within a specific selection of patients, the prevalence of disease and disease severity may be higher. In case of a more advanced stage of disease, the sensitivity will increase when abnormalities are easily detected. The relation between prevalence and clinical setting is illustrated by the comparison between a hospitalized bedridden patient who develops sudden dyspnea and hypoxia accompanied by an elevated level of d-dimers, and an otherwise healthy person who presents with acute dyspnea in an outpatient setting. Clearly, the first patient has a higher probability of pulmonary embolism than the second. Because the first patient was bedridden, more extensive pulmonary embolism could have been developed than in the second patient, which may increase the likelihood that embolism is detected on subsequent pulmonary CT angiography, thus increasing sensitivity. Conversely, in a screening setting, the prevalence of disease is low and disease stages are likely to be less advanced, which may increase the specificity. Hence, test characteristics can vary with the population in which the test is applied. The influence of other factors, such as age, sex, and comorbidity, on test characteristics can be evaluated by multivariate regression analysis [18,19].

Although seemingly straightforward, a considerable variation in study design in test research has been observed. To overcome inconsistent methodology and reporting on test research, the STARD initiative (Standards for Reporting of Diagnostic Accuracy) was launched [20].

## 2.2. Step 2: Added value of a test

In clinical practice, tests are generally used in sequence. For each subsequent test, its added value must be considered. The added value in this phase of evaluation is expressed in terms of increase in proportion of patients correctly categorized as diseased or nondiseased, which is represented by an increased area under the receiver-operating characteristic (AUROC) curve.

Added value of a new test can be estimated using multivariate regression analysis [3,8].

In addition, a statistically significant increase in the AUROC curve does not necessarily represent clinical improvement, because this also depends on the positivity criterion. In patients with clinically suspected carotid artery stenosis, the added value of imaging will not inform us whether clinical outcome will improve in terms of strokes avoided. If MRA enables detecting smaller grade stenosis

than DUS, more patients will have a positive test result. When patients with carotid artery stenosis benefit from treatment of severe stenosis only, which could already be accurately detected by DUS, detection of smaller grade stenosis by MRA will not improve outcome. Thus, added value in terms of a statistically significant increase in AUROC curve may not always have actual clinical value.

### 2.3. Step 3: Clinical outcome

The eventual goal of a new diagnostic test is to improve clinical outcome [3–5,8]. Clinical outcome after implementation of a new diagnostic test can be assessed by an RCT [2–4,7].

RCTs in diagnostic research, however, have several limitations. Firstly, RCTs generally require a long follow-up before consequences of false-positive and false-negative test results become apparent. During this long period of follow-up, new diagnostic techniques may emerge, which may outdate the results of the trial before it is completed. Secondly, large groups of participants are often needed. When the new and the reference tests do not differ much in their ability to detect disease, only the limited subset of subjects with discordant test results carries relevant information. Dependent on the expected difference in disease course after diagnosis and ensuing treatment, the overall difference in outcome between the randomized groups may be further diluted. Therefore, a straightforward comparison of two diagnostic tests may already require many participants. Importantly, often more than two diagnostic tests require simultaneous evaluation. Additionally, various cutoff criteria, including multiple test results, the order of tests, and specific combinations of tests, will increase the number of diagnostic strategies to be evaluated. Representing all diagnostic strategies implies that the number of participants required increases exponentially. In the example of carotid stenosis, test characteristics of DUS, MRA, and DSA show little difference. An RCT should include ample participants to represent all relevant diagnostic strategies and to attain sufficiently large groups with discordant test results. Finally, a diagnostic RCT including an already established treatment strategy is inefficient. If outcome of treatment has already been established by therapeutic research, follow-up of treated patients would be redundant. In carotid stenosis, the effect on clinical outcome after carotid endarterectomy had already been assessed in therapeutic trials [9,10]. Therefore, follow-up after treatment was deemed unnecessary for the evaluation of diagnostic strategies. An RCT would be a viable option only if the new diagnostic tool has consequences for treatment that has not been previously evaluated.

Decision-analytic models present an important alternative to an RCT [2,5]. Through modeling, clinical outcome after application of a new test can be predicted by integrating the available evidence. A decision tree gives an overview of all diagnostic strategies with probabilities for

each event. A decision tree will not suffice when events tend to recur or when events occur after a long and variable time span, such as in chronic diseases [21,22]. For more complex diagnostic and disease processes, such as in carotid artery stenosis, Markov models are useful. These are typically based on probabilities of transitions between predefined health states for all possible scenarios. Time-dependent risks and instantaneous risks for each scenario can be included. The distribution of patients among the different health states after a specified time span, for example, 1 year, is calculated. This can repeatedly be performed depending on the defined duration of follow-up. Uncertainty regarding the estimates of the input parameters of the model can be explored using Monte Carlo simulation. Through Monte Carlo simulation, multivariable sensitivity analysis is conducted to predict clinical outcome of a hypothetical group of subjects after a defined period of time, including uncertainty regarding the predicted outcome [22–24]. Decision-type models are flexible and can be designed to reflect an infinite number of diagnostic strategies. Furthermore, models can simulate a long follow-up period for a large group of subjects while recording all relevant outcomes and time frames simultaneously. Because there is no need to perform a lengthy RCT, the duration of uncertainty regarding the best diagnostic strategy is considerably shortened. Moreover, although decision models may appear complex and modeler driven, they do not obscure clinical evidence. All assumptions and estimates can be verified.

The level of decision uncertainty with regard to the appropriate diagnostic strategy is mostly reduced by an RCT. Decision modeling forms the only alternative where a proper RCT is unfeasible. Uncertainty remaining after a modeling study may be reduced by further research on the parameter that introduces most uncertainty. Whether investment of additional resources to reduce uncertainty is worthwhile can be formally assessed [25].

There are some limitations to decision models. To create the model, sufficient information should be available. Furthermore, diagnostic processes may be too complex to summarize in a decision tree. In addition, decision models may never comprise all subtle associations and interactions of clinical reality. For instance, test characteristics are generally assumed to be constant and independent of other test results acquired previously. In reality, these previous test results can influence the interpretation of the subsequent test as being discussed before, but the actual change in test characteristics is rarely known. By sensitivity analysis, the influence of change in test characteristics on the eventual outcome can be assessed.

### 2.4. Step 4: Cost-effectiveness

Decision models can also keep track of resource use in relation to clinical outcome. The costs of the tests applied, the transitions, and events occurring can be integrated in the model. Subsequently, the incremental costs per additional

unit of health, for example, extra costs per quality-adjusted life-year (QALY) gained, can be calculated for all diagnostic strategies. By comparing these results, the strategy with the best trade-off between costs and effects can be identified.

A study comparing 62 diagnostic strategies for carotid artery stenosis was performed. Using a Markov model, long-term outcome of the diagnostic work-up and following treatments in terms of life-years, QALYs, and costs were predicted and compared across strategies. DUS alone resulted in an optimal balance between costs and effects. The addition of MRA led to a slightly better outcome with a disproportionate increase in costs. Because of its complication risk, any strategy comprising DSA was proven inferior [26].

This demonstrates that Markov modeling not only allows for assessment of cost-effectiveness, but also for evaluation of clinical outcome. Hence, this example covers steps 3 and 4.

### 3. Evaluation of decision models

Examples of validated decision-analytic models showed that the results of decision modeling were consistent with observed data from cohort follow-up studies [27–29]. For two of these studies, the input parameters of the Markov model originated from a cohort different from that with which the results were compared [27,28]. Another study used the short-term follow-up data of an observational study cohort for a Markov model. The expected value of the model was compared with the long-term follow-up data of the same cohort and with data from the literature [29].

A few aspects are important for the evaluation of Markov models. First, the model should reflect clinical reality in sufficient detail such that critical evaluation by expert peers remains possible. Furthermore, all model components, that is, the parameters describing each state transition, the health state value judgments, and cost estimates, need to be discussed in a multidisciplinary setting. Finally, because Markov models will be increasingly used, guidelines for its evaluation need to be developed [30,31].

We strongly recommend using decision models to evaluate diagnostic tests, provided reliable model parameters can be obtained, particularly, estimates of key variables. Notably, a reference standard is required to evaluate any diagnostic approach, because it remains essential for the diagnosis of any disease. This reference standard does not necessarily have to be a “gold” standard. Even though imperfection of a reference standard impairs perfect labeling of “false”-positive and “false”-negative test results, the impact of uncertainty of test characteristics on clinical outcome can be evaluated by Monte Carlo simulation. Finally, a standard treatment and pertaining outcome should be established. This can range from supportive treatment to causal treatment. For most of the diseases, a certain treatment can be recognized. Hence, we believe that, in most cases, a decision model has major advantages compared with diagnostic RCTs, because it saves time and resources, and allows cost-effectiveness analysis.

### 4. Conclusions

Evaluation of a diagnostic strategy requires a phased approach. For assessment of outcome in addition to test characteristics frequently, an RCT is neither feasible nor warranted. Decision-analytic models can integrate the best available evidence, including economic data, and should be part of the standard methodology in diagnostic research.

### Acknowledgments

The funding for the study was provided by Netherlands Organization for Scientific Research (NWO), the Netherlands Organization for Health Research and Development (ZonMw), 945-04-310.

Conflicts of interests for all authors: none declared.

### References

- [1] Feinstein AR. Misguided efforts and future challenges for research on “diagnostic tests”. *J Epidemiol Community Health* 2002;56:330–2.
- [2] Hunink MG, Krestin GP. Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology. *Radiology* 2002;222:604–14.
- [3] Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. *BMJ* 2002;324:477–80.
- [4] Glud C, Glud LL. Evidence based diagnostics. *BMJ* 2005;330:724–6.
- [5] Mol BW, Lijmer JG, Evers JL, Bossuyt PM. Characteristics of good diagnostic studies. *Semin Reprod Med* 2003;21:17–25.
- [6] Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ* 2002;324:539–41.
- [7] Van den Bruel A, Cleemput I, Aertgeerts B, Ramaekers D, Buntinx F. The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. *J Clin Epidemiol* 2007;60:1116–22.
- [8] Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem* 2004;50:473–6.
- [9] European Carotid Surgery Trialists’ Collaborative Group. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). *Lancet* 1998;351:1379–87.
- [10] Barnett HJ, Taylor DW, Eliasziw M, Fox AJ, Ferguson GG, Haynes RB, et al. Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. *N Engl J Med* 1998;339:1415–25.
- [11] Nederkoorn PJ, Van Der Graaf Y, Hunink MG. Duplex ultrasound and magnetic resonance angiography compared with digital subtraction angiography in carotid artery stenosis: a systematic review. *Stroke* 2003;34:1324–32.
- [12] Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol* 2003;56:1118–28.
- [13] Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061–6.
- [14] Nederkoorn PJ, Mali WP, Eikelboom BC, Elgersma OE, Buskens E, Hunink MG, et al. Preoperative diagnosis of carotid artery stenosis: accuracy of noninvasive testing. *Stroke* 2002;33:2003–8.
- [15] Gallagher EJ. Clinical utility of likelihood ratios. *Ann Emerg Med* 1998;31:391–7.
- [16] Richardson WS, Wilson MC, Guyatt GH, Cook DJ, Nishikawa J. Users’ guides to the medical literature: XV. How to use an article



- about disease probability for differential diagnosis. Evidence-Based Medicine Working Group. *JAMA* 1999;281:1214–9.
- [17] Knottnerus JA, Leffers P. The influence of referral patterns on the characteristics of diagnostic tests. *J Clin Epidemiol* 1992;45:1143–54.
- [18] Brenner H, Sturmer T, Gefeller O. The need for expanding and re-focusing of statistical approaches in diagnostic research. *J Epidemiol Community Health* 2002;56:338–9.
- [19] Knottnerus JA. Application of logistic regression to the analysis of diagnostic data: exact modeling of a probability tree of multiple binary variables. *Med Decis Making* 1992;12:93–108.
- [20] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin Radiol* 2003;58:575–80.
- [21] Hunink MG, Glasziou PP, Siegel J, Weeks J, Pliskin J, Elstein A, et al. Recurring events. In: Hunink MG, editor. *Decision making in health and medicine—integrating evidence and values*. Cambridge: Cambridge University Press; 2001. p. 305–38.
- [22] Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decis Making* 1993;13:322–38.
- [23] Beck JR, Kassirer JP, Pauker SG. A convenient approximation of life expectancy (the “DEALE”). I. Validation of the method. *Am J Med* 1982;73:883–8.
- [24] Hunink MG, Glasziou PP, Siegel J, Weeks J, Pliskin J, Elstein A, et al. Variability and uncertainty. In: Hunink MG, editor. *Decision making in health and medicine—integrating evidence and values*. Cambridge: Cambridge University Press; 2001. p. 339–63.
- [25] Claxton K, Cohen JT, Neumann PJ. When is evidence sufficient? *Health Aff (Millwood)* 2005;24:93–101.
- [26] Buskens E, Nederkoorn PJ, Buijs-Van Der Woude T, Mali WP, Kappelle LJ, Eikelboom BC, et al. Imaging of carotid arteries in symptomatic patients: cost-effectiveness of diagnostic strategies. *Radiology* 2004;233:101–12.
- [27] Martin C, Vanderpump M, French J. Description and validation of a Markov model of survival for individuals free of cardiovascular disease that uses Framingham risk factors. *BMC Med Inform Decis Mak* 2004;4:6.
- [28] Nijhuis RL, Stijnen T, Peeters A, Witteman JC, Hofman A, Hunink MG. Apparent and internal validity of a Monte Carlo-Markov model for cardiovascular disease in a cohort follow-up study. *Med Decis Making* 2006;26:134–44.
- [29] Welsing PM, Severens JL, Hartman M, van Gestel AM, van Riel PL, Laan RF. The initial validation of a Markov model for the economic evaluation of (new) treatments for rheumatoid arthritis. *Pharmacoeconomics* 2006;24:1011–20.
- [30] Krahm MD, Naglie G, Naimark D, Redelmeier DA, Detsky AS. Primer on medical decision analysis: Part 4—analyzing the model and interpreting the results. *Med Decis Making* 1997;17:142–51.
- [31] Richardson WS, Detsky AS. Users’ guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1995;273:1292–5.